Part I

Introduction

Chapter 1

Made in The World

It does not feel like a long time ago when I began my undergraduate studies in Economics at Universitat Pompeu Fabra (UPF), the same institution that hosts the Centre de Recerca en Economia Internacional (CREI), where these Lectures were delivered. It was 1994 and I felt I lived in a truly global economy. The music I listened to and the movies I watched were mostly British or American. Most of the clothes I wore were manufactured abroad, some of them in rather exotic places such as Morocco or Taiwan. My favorite beer was Dutch. At UPF, about half of my teachers were foreign, a third of the classes were taught in English and most of the textbooks were the same ones used in universities around the globe.

In hindsight, it seems pretty clear, however, that the world had not yet witnessed the full advent of globalization. What has changed since 1994? *First* and foremost, the last two decades have brought a genuine information and communication technology (ICT) revolution that has led to a profound socioeconomic transformation of the world in which we live. The processing power and memory capacity of computers have doubled approximately every two years (as implied by Moore's law), while the cost of transmitting a bit of information over an optical network has decreased by half roughly every nine months (a phenomenon often referred to as Butter's law). The number of internet users has increased by a factor of 100, growing from around 25 million users in 1994 to more than 2,500 million users in 2012 (see World Development Indicators). As a result of these technological developments, the cost of processing and transmitting information at long distances has dramatically fallen in recent times. Consider the following example: in 2012, the 3.3GB file containing my favorite movie of 1994, *Pulp Fiction*, could be downloaded from Amazon.com in about 11 minutes and 16 seconds using a standard broadband connection with a download speed of 5 megabits per

second. In 1994, downloading that same file using a dial-up connection and the state-of-the-art modem, which allowed for a maximum speed of 28.8 kilobits per second, would have kept your phone line busy for at least 33 hours and 23 minutes![1]

Second, during the same period, governments have continued (and arguably intensified) their efforts to gradually dismantle all man-made trade barriers. This process dates back to the initial signing of the General Agreement on Tariffs and Trade (GATT) in 1947, but it has experienced a revival in the 1990s and 2000s with the gradual expansion of the European Union, the formation of the North American, Mercosur, and ASEAN free trade agreements, the signing of a multitude of smaller preferential trade agreements under the umbrella of GATT's Article XXIV, and China's accession to the World Trade Organization (WTO), just to name a few. As a consequence, the world's weighted average tariff applied on traded manufactured goods fell from 5.14% in 1996 to 3.03% in 2010 (see World Development Indicators).[2]

Third, political developments in the world have brought about a remarkable increase in the share of world population actively participating in the process of globalization. These changes largely stemmed from the fall of communism in Eastern Europe and the former Soviet Union, but also from an ensuing ideological shift to the right in large parts of the globe. Thus, not only did former communist countries embrace mainstream capitalist policies, but these policies themselves became more friendly towards globalization, as exemplified by the deepening of trade liberalization mentioned in the last paragraph, but also by a notable relaxation of currency convertibility and balance of payments restrictions in several low and middle-income countries.[3]

The Slicing of the Value Chain

One of the manifestations of these three developments in the world economy has been a gradual disintegration of production processes across borders.

[1] Paraphrasing a memorable quote from Samuel L. Jackson's character in *Pulp Fiction*, download speeds today and in 1994 "ain't the same [*freaking*] ballpark. They ain't the same league. They ain't even the same [*freaking*] sport."

[2] Technological developments since 1994 have also reduced the quality- (or time-) adjusted costs of transporting goods across countries (see Hummels, 2007), while investments in infrastructure in less developed economies have also contributed to spreading the effects of globalization across regions in those countries.

[3] The late 1990s also saw the emergence of a left-leaning anti-globalization movement, which drew particular attention during the 1999 WTO meetings in Seattle. There is little evidence, however, of this movement having led to any significant slow down in the process of globalization (see, for instance, Harrison and Scorse, 2010).

More and more firms now organize production on a global scale and choose to offshore parts, components or services to producers in foreign and often distant countries. The typical "Made in" labels in manufactured goods have become archaic symbols of an old era. These days, most goods are "Made in the World."

A variety of terms have been used to refer to this phenomenon: the "slicing of the value chain", "fragmentation of the production process", "disintegration of production", "delocalization", "vertical specialization", "global production sharing", "unbundling", "offshoring" and many more (see Feenstra, 1998). I shall use these terms interchangeably throughout the book.[4]

The case of Apple's iPad 3 tablet nicely illustrates the magnitude of this new form of globalization. The slim and sleek exterior of the tablet hides a complex manufacturing process combining designs and components provided by multiple suppliers with operations in various countries. Although Apple does not disclose detailed information on its input providers, a clear picture of the global nature of the iPad 3 production process emerges when combining information from tear-down reports (such as those published by isuppli.com and ifixit.com) with various press releases.[5] For instance, it is well-known that the tablet itself is assembled in China (and since 2012 also in Brazil) by two Taiwan-based companies, Foxconn and Pegatron. The revolutionary retina display is believed to be manufactured by Samsung of South Korea in its production plant in Wujiang City, China. The distinctive touch panel is produced (at least, in part) by Wintek, a Taiwan-based company that also owns plants in China, India and Vietnam, while the case is provided by another Taiwanese company, Catcher Technologies, with operations in Taiwan and China. A third important component, the battery pack, also originates in Taiwan and is sold by Simplo Technologies and Dynapack International. Apart from these easily identifiable parts, the iPad 3 incorporates a variety of chips and other small technical components provided by various

[4] At times, I will also use the buzzword "outsourcing", but I will do so only when referring to *arm's-length* sourcing relationships, that is instances of fragmentation in which the firms exchanging parts are not related (i.e., integrated). Outsourcing is often observed not only in foreign but also in domestic vertical relationships.

[5] Facing strong criticism over the working conditions in its suppliers' factories, Apple released a full list of its 156 global suppliers early in 2012 (see http://images.apple.com/supplierresponsibility/pdf/Apple_Supplier_List_2011.pdf). Teardown reports further faciliate a mapping between the iPad parts and their respective producers. Press releases sometimes also identify particular suppliers with specific iPad 3 components (see, for instance, Forbes' "Batteries Required?" available at http://www.forbes.com/global/2010/0607/best-under-billion-10-raymond-sung-simplo-technology-batteries-requried.html).

firms with headquarters and R&D centers in developed economies and manufacturing plants scattered around the world. A non-exhaustive list includes (again) Korea's Samsung, which is believed to manufacture the main processor (designed by Apple), U.S.-based Qualcomm supplying 4G modules, and Italo-French STMicroelectronics contributing key sensors.[6]

Apple's sourcing strategies are far from being an isolated example of a global approach to the organization of production. In fact, the increasing international disintegration of production processes has been large enough to be salient in aggregate statistics. During the 1990s and early 2000s, when this phenomenon was still in its infancy, researchers devised several approaches to measuring the quantitative importance of global production sharing.[7] Feenstra and Hanson (1996*b*), for instance, used U.S. Input-Output tables to infer the share of imported inputs in the overall intermediate input purchases of U.S. firms; they found that this share had already increased from 5.3% in 1972 to 11.6% in 1990. Campa and Goldberg (1997) found similar evidence for Canada and U.K., but surprisingly not for Japan, where the reliance on foreign inputs appeared to have declined between 1974 and 1993. Hummels, Ishii and Yi (2001) instead constructed a measure of vertical specialization capturing the value of imported intermediate inputs (goods and services) embodied in a country's exported goods and found that it already accounted for up to 30% of world exports in 1995, having grown by as much as 40% since 1970.

The work of Johnson and Noguera (2012*a*, 2012*b*) constitutes the state of the art in the use of Input-Output tables to quantify the importance of global production sharing and its evolution in recent years. The main innovation of their methodology is in the attempt to compute a *global* Input-Output table from which one can back out the value-added and intermediate input contents of gross international trade flows. In particular, their VAX ratio (the value-added to gross-value ratio of exports) is an appealing inverse measure of the importance of vertical specialization in the world production: the lower is this measure, the larger is the value of imported inputs embodied in exports.[8] As is clear from Figure 1.1, their VAX ratio has declined rather significantly

[6] A more extensive list can be found at: http://www.chipworks.com/en/technical-competitive-analysis/resources/blog/the-new-ipad-a-closer-look-inside/.

[7] The task is complicated by the fact that data on trade flows of goods is collected on a gross output basis, without regard to the particular sources of the value added embodied in these goods.

[8] In a very recent paper, Koopman, Wang and Wei (2014) devise a methodology that nicely nests Johnson and Noguera's (2012) VAX measure with the vertical specialization measures developed by Hummels, Ishii and Yi (2001).

since 1970 with about two-thirds of the decline occurring after 1990. Johnson and Noguera (2012*b*) show that this decline is explained solely by increased offshoring within manufacturing. Furthermore, they also find that global production sharing has grown disproportionately in emerging economies and also appears to increase following the signing of regional trade agreements.

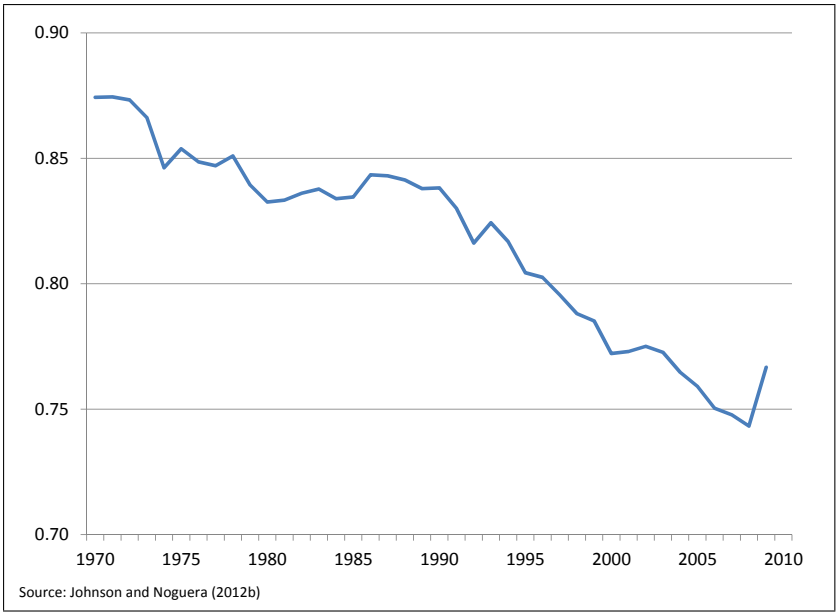

Figure 1.1: Ratio of Value Added to Gross Exports (VAX), 1970-2009

Two limitations of the fragmentation measures discussed so far are that they rely on fairly aggregated Input-Output data and that they impose strong proportionality assumptions to back out the intermediate input component of trade. A different approach to measuring the degree to which production processes are fragmented across countries was first suggested by Yeats (2001), and consists of computing the share of trade flows accounted for by SITC Rev.2 industry categories that can be safely assumed to contain *only* intermediate inputs (as reflected by the use of the word "Parts of" at the beginning of the category description). It turns out that all these industries are in the "Machinery and Transport Equipment" industrial group (or SITC 7). Yeats (2001) found that intermediate input categories accounted for about 30% of OECD merchandise exports of machinery and transport equipment in 1995, and that this share had steadily increased from its 26.1% value in 1978. A limitation of Yeats' measure is that, by focusing on industries composed *exclusively* of inputs, it naturally understates the importance of input trade.

This might explain why when updating this methodology to present times, one finds little evidence of a further increase in this share.[9]

An alternative to categorizing trade flows as *either* final goods or intermediate inputs is to attempt to calculate a more continuous measure of the "upstreamness" of the goods being traded. This is the approach in Antràs, Chor, Fally and Hillberry (2012), who use Input-Output data to construct a weighted index of the average position in the value chain at which an industry's output is used (i.e., as final consumption, as direct input to other industries, as direct input to industries serving as direct inputs to other industries, and so on), with the weights being given by the ratio of the use of that industry's output in that position relative to the total output of that industry. Intuitively, the higher this measure is, the more removed from final good use (and thus the more upstream) is that industry's output. The Data Appendix contains a lengthier discussion of the construction of this index.[10] Antràs, Chor, Fally and Hillberry (2012) use the measure to characterize the average upstreamness of exports of different countries in 2002, but it can also be employed to illustrate how the upstreamness of world exports has evolved in recent years. As shown in Figure 1.2, world exports became significantly more upstream in recent years, particularly in the period 2002-08. The patterns are in line with those illustrated in Figure 1.1, and also suggest an increasing predominance of input trade in world trade. Although a significant share of the observed increase in upstreamness is related to an increase the relative weight of petroleum-related industries, even when netting those out, one observes a significant upward trend in upstreamness (see Figure 1.2). Interestingly, both Figures 1.1 and 1.2 identify a disproportionate fall in global production sharing relative to the overall fall in world trade during the early years of the recent 'great recession'.

[9] Other authors attempting to compute the share of intermediate inputs in world trade using alternative methodologies have also found little evidence of a trend in the series (see, for instance, Chen, Kondratowicz and Yi, 2005, or Miroudot, Lanz and Ragoussis, 2009). I have obtained similar results when computing the relative growth of overall trade and input trade using the classification of goods developed by Wright (2014). As argued by Johnson and Noguera (2012*b*), even when taking this finding at face value, it is not necessarily inconsistent with the observed rise in indices of vertical specialization, which better capture the use of imported inputs in producing goods *that are exported.*

[10] This upstreamness index was independently developed by Antràs and Chor (2013) and Fally (2012), and its properties were further studied in Antràs, Chor, Fally and Hillberry (2012).

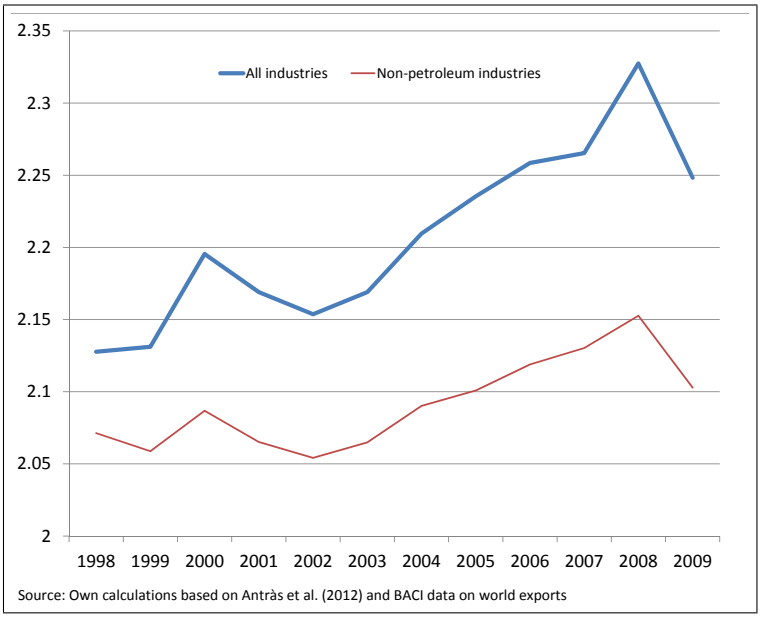

Figure 1.2: Average Upstreamness of World Exports, 1998-2009

Old and New Theories

The noticeable expansion in input trade has also captured the attention of international trade theorists eager to bridge the apparent gap between the new characteristics of international trade in the data and the standard representation of these trade flows in terms of final goods in traditional and new trade theory.

One branch of this new literature has focused on incorporating the notion of fragmentation in otherwise neoclassical models with homogeneous goods, perfectly competitive markets and frictionless contracting. Key contributions include Feenstra and Hanson (1996*a*), Jones (2000), Deardorff (2001), and Grossman and Rossi-Hansberg (2008). The main idea in these contributions is that the production process (as represented by an abstract mapping between factors of production and final output) can be decomposed into smaller parts or stages that are themselves (partly) tradable. Different authors assign different labels to these parts: some refer to them as intermediate inputs, others call them vertical production stages, while others view them as tasks. Regardless of the interpretation of the process under study, a common lesson from this body of work is that the possibility of fragmentation generates nontrivial effects on productivity, and that these endogenous changes in productivity in turn deliver novel predictions for the effects of reductions in

trade costs on patterns of specialization and factor prices. Antràs and Rossi-Hansberg (2009) elaborate on this broad interpretation of this branch of the literature and also offer more details on the specific results of each of these contributions.[11]

As insightful as this body of work has proven to be, it seems clear that modeling global production sharing as simply an increase in the tradability of homogeneous inputs across countries misses important characteristics of intermediate input trade. Prominent among these features is the fact that parts and components are frequently customized to the needs of their intended buyers (remember our example above with the iPad 3). In other words, the disintegration of the production process is more suitably associated with the growth of trade in *differentiated* (rather than homogeneous) intermediate inputs.[12]

Another important characteristic of global production networks is that they necessarily entail intensive *contracting* between parties located in different countries and thus subject to distinct legal systems. In a world with perfect (or complete) contracting across borders, this of course would be of little relevance. Unfortunately, this is not the world we (or at least, I) live in. Real-world commercial contracts are incomplete in the sense that they cannot possibly specify a course of action for *any* contingency that could arise during the course of a business relationship. Of course, the same can be said about domestic commercial transactions, but the cross-border exchange of goods cannot generally be governed by the same contractual safeguards that typically accompany similar exchanges occurring within borders.

Given the subject of this book, it is worth pausing to describe in more detail some of the factors that make international contract enforcement particularly problematic.

[11] Another common feature of the theoretical frameworks developed in these papers is that the number of primitive factors of production is assumed to be small, and normally equal to two. Another branch of the literature has developed perfectly-competitive, frictionless models in which offshoring results from the assignment of a population of a large number of heterogeneous agents into international hierarchical teams (see Kremer and Maskin, 2006, or Antràs, Garicano and Rossi-Hansberg, 2006).

[12] Admittedly, there does not exist much evidence to substantiate this claim. Antràs and Staiger (2012*a*) offer a back-of-the envelope quantification applying the methodology suggested by Schott (2004) to identify international trade in intermediate goods and using the "liberal" classification of Rauch (1999) to distinguish between differentiated and homogeneous goods. They find that the share of differentiated inputs in world trade more than doubled between 1962 and 2000, increasing from 10.56% to 24.85% of world trade. Behar and Freund (2011) show that during the late 1990s and 2000s, intermediate inputs traded within the EU became more sophisticated and involved more relationship-specific investments (in the sense of Nunn, 2007).

Contracts in International Trade

A first natural difficulty in contractual disputes involving international transactions is determining which country's laws are applicable to the contract being signed. In principle, the parties can include a *choice-of-law* clause specifying that any dispute arising under the contract is to be determined in accordance with the law of a particular jurisdiction, regardless of where that dispute is litigated. Nevertheless, many international contracts do not include that clause and, in any case, it is up to the court of law adjudicating a dispute to decide whether it will uphold the expressed desire of the parties. If the court is not familiar with the law specified in the contract, as may often occur in international transactions, the court might decide to rule on the basis of its own law, or they may inadvertently apply the desired foreign law incorrectly.

A second difficulty relates to the fact that even when local courts are competent (in a legal sense), judges may be reluctant to rule with regard to a contract dispute involving residents of foreign countries, especially if such a ruling would entail an unfavorable outcome for local residents. The evidence on the home bias of local courts is mixed, but even those authors advocating that a formal analysis of case law does not support the hypothesis of biases against foreigners readily admit a widespread belief of the existence of such xenophobic biases (see Clermont and Eisenberg, 2007).

A third complication with international contracts relates to the enforcement of remedies stipulated in the court's verdict. For instance, the court might rule in favor of a local importer that was unsatisfied with the quality of certain components obtained from an exporter, and the verdict might require the exporter to compensate the importer for any amount already paid for the components, as well as for any court or even attorney fees incurred. An issue arises, however, if the exporter does not have any assets (say bank accounts or fixed assets) in the importer's country. In that case, it is not clear that the exporter will feel compelled to accept the verdict and pay the importer.

In recent years, there have been several coordinated attempts to reduce the contractual uncertainties and ambiguities associated with international transactions. A particularly noteworthy example is the United Nations Convention on Contracts for the International Sale of Goods (or CISG), or Vienna Convention, which attempts to provide a set of uniform rules to govern contracts for the international sale of goods. The idea is that even when an international contract does not include a choice-of-law clause, parties whose places of business are in different signing countries can rely on the Convention

to protect their interests in courts. As ambitious as the CISG initiative is, it has arguably fallen short of its objectives. For instance, several countries or regions (most notably, Brazil, Hong Kong, India, South Africa, Taiwan, and the United Kingdom) have yet to sign the agreement. Furthermore, a few of the signing countries have expressed reservations and choose not to apply certain parts of the agreement. Finally, it is not uncommon for private parties to explicitly opt out of the application of the Convention, as allowed by its Article 6. The reluctance to unreservedly embrace the Convention has been associated with the somewhat vague language of the text, which might foster the natural inclination of judges to interpret the Convention through the lens of the laws of their own State.[13]

Another attempt to ameliorate the perceived contractibility of international transactions consists in resorting to international arbitration. More specifically, an international trade contract can include a (so-called) *forum-of-law* clause establishing that a particular arbitrator, such as the International Chamber of Commerce (ICC) in Paris, will resolve any contractual dispute that may arise between the parties. International arbitration is appealing because it avoids the aforementioned uncertainties associated with litigation in national courts. It is also relatively quick and parties benefit from the fact that arbitrators tend to have more commercial expertise than a typical judge. Furthermore, arbitration rulings are confidential and are generally perceived to be more enforceable than those of national courts because they are protected by the Convention on the Recognition and Enforcement of Foreign Arbitral Awards, also known as New York Convention. Despite its attractive features, international arbitration is rarely used in practice because its cost is too high for most firms to bear.[14]

One might argue that even when explicit contracts are incomplete and

[13] The Institute of International Commercial Law at Pace Law School maintains a website (http://www.cisg.law.pace.edu/) with comprehensive information on the CISG, including a database of thousands of legal cases in which the Convention was invoked. The details of these cases offer a vivid account of the nature of contractual disagreements in international trade.

[14] It may be instructive to illustrate this claim with some figures. Using the arbitration cost calculator available from the ICC website, the estimated cost of arbitration (involving a single arbitrator) would be $5,401 for a $10,000 dispute (or a 54% cost-to-dispute-amount ratio), $15,425 for a $100,000 dispute, $61,094 for a $1 million dispute, and $170,799 for a $10 million dispute (or a mere 1.7% cost-to-dispute-amount ratio). It is thus little surprise that there were only 796 ICC arbitration requests in 2011 and that the amount in dispute was under one million U.S. dollars in only 22.7% of these cases (see http://www.iccwbo.org/products-and-services/arbitration-and-adr/arbitration/cost-and-payment/cost-calculator/).

perceived to be unenforceable, parties in international transactions can still resort to implicit contracting to sustain 'cooperation'. We shall briefly develop this idea in Chapter 3. Nevertheless, it is particularly difficult to render international commercial relationships self-enforcing. On the one hand, international parties are less likely to meet face-to-face and to transact on a repeated basis than domestic parties, in part due to distance and trade costs, but also due to shocks (such as exchange rate movements) that can quickly turn efficient relationships into inefficient ones. On the other hand, the possibility of collective or community enforcement is hampered again by distance but also by the fact that parties might have different cultural and societal values. In sum, and in the words of Rodrik (2000), "ultimately, [international] contracts are often neither explicit nor implicit; they simply remain incomplete."

Although contractual risks are also of relevance for the exchange of final goods (see Chapter 3), the detrimental effects of imperfect international contract enforcement are likely to be particularly acute for transactions involving intermediate inputs. This is so for at least two reasons. First, input transactions are often associated with relatively long time lags between the time an order is placed (and the contract is signed) and the time the goods or services are delivered (and the contract is executed). Second, parts and components often entail significant relationship-specific investments and other sources of lock-in on the part of both buyers and suppliers, which make contractual breaches particularly costly. As argued above, suppliers often customize their output to the needs of particular buyers and would find it difficult to sell those goods to alternative buyers, should the intended buyer decide not to abide by the terms of the contract. Similarly, buyers often undertake significant investments whose return can be severely diminished by incompatibilities, production line delays or quality debasements associated with suppliers not going through with their contractual obligations.[15]

Firm Responses to Contractual Insecurity

When designing their global sourcing strategies, firms face two key decisions. The first one concerns the *location* of the different stages in the value chain

[15] A third more specific reason for which input trade might be perceived to be less contractually secure relates to the fact that Article 3 of CISG explicitly excludes from the applicability of the Convention situations in which "the party who orders the goods undertakes to supply a substantial part of the materials necessary for such manufacture or production," thus making the Convention less relevant for sustaining cooperation in global production sharing networks.

and involves deciding in which country or region firms will conduct R&D and product development, where parts and components should be produced, what is the best place to assemble the finished good, and so on. The second key decision relates to the extent of *control* that firms exert over these different production stages. For instance, firms may decide to keep these production stages within firm boundaries, thus engaging in foreign direct investment (FDI) when the integrated entity is in a foreign country. Other firms may be less inclined to keep a tight control over certain stages and thus choose to contract with suppliers or assemblers at arm's-length.

Neoclassical models of fragmentation focus exclusively on the first of these decisions and emphasize that fragmentation will emerge as part of a competitive equilibrium whenever firms find it cost-minimizing to break up production processes across countries. The source of the cost-advantage associated with fragmentation varies by model; sometimes it stems from differences in relative factor endowments across countries (which, for instance, naturally confer comparative advantage in labor-intensive stages to relatively labor-abundant countries), while other times they are motivated by technological differences across countries.

Neoclassical models are silent on the issue of control. This is not because these models assume perfect competition, constant returns to scale, or homogeneous goods. Instead, the key assumption that renders those models (and just about *any* model in the field of International Trade) vacuous when tackling the notion of control is the assumption of perfect or complete contracting. Indeed, if firms could foresee all possible future contingencies, and if they could costlessly write contracts that specify in an enforceable manner the course of action to be taken in all of these possible contingencies, then firms would no longer need to worry about "controlling" the workers, the internal divisions or the supplying firms with whom they interact in production. The complete contract would in fact confer *full* control to the firm regardless of the ownership structure that governs the transactions between all these producers. In other words, and as Coase (1937) anticipated more than seventy-five years ago, firm boundaries are indeterminate in a world of complete contracts.[16]

In the real world, however, contracts are very much incomplete and especially so in international transactions, where as argued above, the enforce-

[16]It is worth stressing that even in the presence of product differentation and market power, firm boundaries remain indeterminate when contracts are complete. For example, the often-cited double-marginalization rationale for vertical integration rests on the assumption that firms and suppliers cannot sign simple two-part tariff contracts, and as such, it also constitutes an incomplete-contracting theory of firm boundaries.

ability of contracts is particularly questionable. In response to this perceived contractual insecurity, firms spend a substantial amount of time and resources figuring out the best possible way to *organize* production in the global economy. In some cases, foreseeing that producers located in a particular country might not feel compelled to follow through with their contractual obligations, firms contemplating doing business in that country might decide to do so within their firm boundaries, either by setting up a new, wholly- or partially-owned affiliate or by acquiring a controlling stake in an existing firm in that country. In some circumstances, however, the lack of contract enforceability might precisely turn firms to independent suppliers for the procurement of parts because such an arrangement might elicit the best performance from foreign producers. In other words, it is important to keep in mind that internalization is a double-edged sword: it may partly protect the integrating party from the vagaries of international contracting, but it might dilute the incentives to produce efficiently of the integrated party, which is now more tightly controlled and has less power in the relationship (cf., Grossman and Hart, 1986).

The boundaries of firms in the world economy are thus the result of the (constrained) optimal decisions of firms attempting to organize production in the most profitable way possible. A recurring theme of this book, particularly in Part III, is that much can be learned from a theoretical and empirical study of the *fundamental* forces that appear to shape whether international transactions are internalized or not, independently of the firm or sector one is studying.

Some readers might be asking themselves at this point: why should one care about the boundaries of multinational firms? Surely, the fact that we can write testable models of the internalization decision is not a convincing enough argument to care about it. A first answer to this question is that understanding the boundaries of firms, and of multinational firms in particular, *is* interesting in its own right. Ever since the pioneering work of Ronald Coase (1937), this topic has preoccupied the minds of many distinguished economists, and constitutes one of the central themes of the field of Organizational Economics. A second, perhaps more compelling answer is that delineating the boundaries of multinational firms constitutes a necessary first stage for properly studying the causal implications of multinational activity on various objects of interest, such as measures of economic activity and growth, absolute and relative factor price movements, and welfare. In other words, because multinational activity is *not* randomly assigned across countries and sectors, understanding the key drivers behind such selection into multinational activity may be crucial for identification purposes. I will

fall short of demonstrating this point in the current book, but I do hope that the stylized models overviewed in Part III will prove to be useful for that purpose.

Practitioners (and perhaps some academics too) might in turn react skeptically to the idea that low-dimensional models can possibly capture the reasoning behind the complex and idiosyncratic decisions of firms in the world economy. Business school cases often highlight the peculiarities of particular organizational decisions, making it hard to envision that much can be gained from extrapolating from those particular cases. The fact that comprehensive datasets on the integration decisions of firms are not readily available might have only compounded this belief, as most empirical studies of integration decisions rely on data from specific industries or firms.[17]

A Comparative Advantage of Trade Statistics

An advantage of studying the global integration decisions of firms is that data on international transactions are particularly accessible due to the widespread existence of official records of goods and services crossing borders. For instance, it is well-known that researchers can easily access data on U.S. imports from any country of the world at the remarkably detailed ten-digit Harmonized Tariff Schedule classification system, which consists of nearly 17,000 categories.[18] A less well-known fact is that, in some countries, these same detailed country- and product-level data contain information on the extent to which trade flows involve related parties or non-related parties. Most notably, the "U.S. Related-Party Trade" data collected by the U.S. Bureau of Customs and Border Protection and managed by the U.S. Census Bureau provides data on related and non-related-party U.S. imports and exports at the six-digit Harmonized System (HS) classification (which consists of over 5,000 categories) and at the origin/destination country level. This amounts to hundreds of thousands of observations *per year* on the relative prevalence of integration versus nonintegration across products and countries.[19]

[17] See Baker and Hubbard (2003) for a particularly careful study using data from the trucking industry, and Lafontaine and Slade (2007) and Bresnahan and Levin (2012) for broad surveys of the empirical literature on vertical integration.

[18] Downloading these data from the NBER website, one can readily verify that in 2001 France exported $15,747 worth of frozen potatoes to the United States (HTS code 2004.10), yet none of those were French fries (HTS code 2004.10.8020)!

[19] The U.S. Related Party Trade data are publicly available at: http://sasweb.ssd.census.gov/relatedparty/. This website permits downloading the data at the six-digit NAICS level. The finer six-digit Harmonized System (HS) data are available from the U.S. Census for a fee, but I have also made them available at

What do these data tell us about the global sourcing strategies of firms? The first thing that one notices when using U.S. related-party trade data is how predominant intrafirm transactions are in U.S. trade. In 2011, intrafirm imports of goods totaled $1,056.2 billion and constituted a remarkable 48.3 percent of total U.S. imports of goods ($2,186.9 billion). In fact, the share of intrafirm trade has been higher than 46.5 percent in every year since 2000. On the export side, related-party exports are also pervasive, with their share in total U.S. exports ranging from 28 to 31 percent in recent years. These figures illustrate the importance of multinational firms for U.S. trade.[20]

A second evident feature of the data is that the share of U.S. intrafirm imports varies widely across countries. On the one hand, in 2011 intrafirm imports equalled 0 for 10 countries and territories (including Cuba), all exporting very low volumes to the U.S. On the other hand, in that same year the share of intrafirm trade reached a record 89.6 percent for U.S. imports from Western Sahara. Leaving aside communist dictatorships and disputed territories, and focusing on the 50 largest exporters to the U.S., Figure 1.3 illustrates that the share of intrafirm trade still varies significantly across countries, ranging from a mere 2.4 percent for Bangladesh to an astonishing 88.5 percent for Ireland.

Similarly, the share of intrafirm trade varies widely depending on the type of product being imported. Again, the raw data contain infrequently traded goods with shares equal to 0 and 100, but even when focusing on the top 20 six-digit HS manufacturing industries by importing volume, in Figure 1.4 one observes significant variation in the share of intrafirm trade, which ranges from a share of 11.4 percent for U.S. imports of sweaters, pullovers and sweatshirts made of cotton (HS 611020) to 98.8 percent for imports of automobiles with engines of more than 3000 cc (HS 870324). This variation persists even when focusing on much narrower sectors. As shown in Figure 1.5, when analyzing imports across subcategories of the four-digit Harmonized System sector 8708 ('Parts and accessories of motor vehicles'), the share of intrafirm trade still ranges from 19.8 percent for drive axles (HS

---

http://scholar.harvard.edu/antras/books.

[20] In contrast, Atalay, Hortacsu and Syverson (2013) study intrafirm shipments across U.S. multiplant firms and find that these constitute a very small share of total shipments, a finding that they interpret as indicating that firm boundaries are shaped by issues related to the the transfer of intangible inputs, rather than of physical goods. However, as argued above, contractual insecurity in the exchange of physical inputs is much more significant in international transactions than in domestic ones, and thus firm boundaries might well be shaped by different factors in cross-border relationships than in the domestic ones in the Atalay, Hortacsu and Syverson (2013) database.

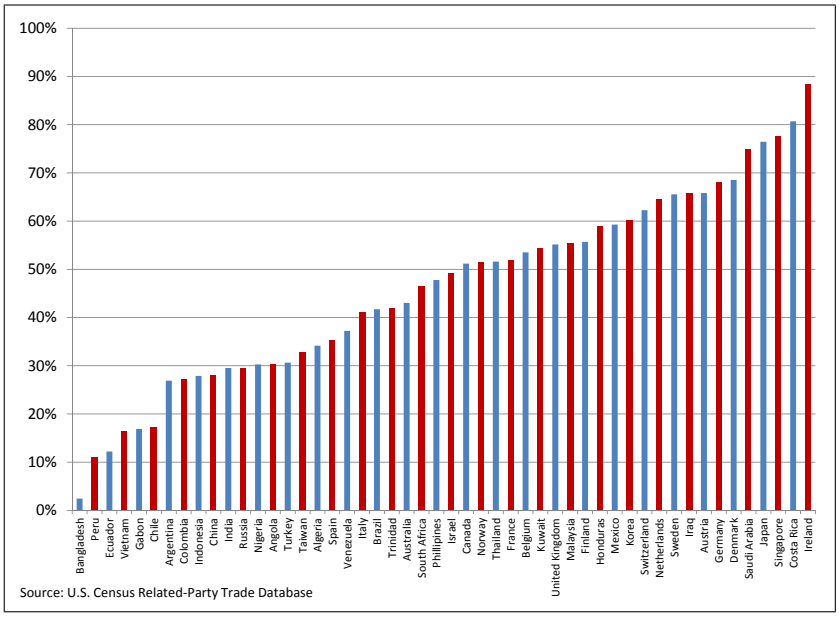

Figure 1.3: Share of U.S. Intrafirm Imports for Largest 50 Exporters to the U.S. in 2011

870850) to 71.2 percent for steering wheels (HS 870894). It is thus clear that U.S.-based producers appear to source different auto parts under quite different ownership structures.

As a final illustration of the richness and variation in the data, consider the six-digit HS industry with the largest share of intrafirm imports in Figure 1.5, namely HS 870894 (Steering Wheels, columns and boxes for motor vehicles). Figure 1.6 reports the share of intrafirm trade for all 56 countries with positive exports to the U.S. in that sector. As is clear from the graph, even when focusing on a narrowly-defined component, a similar pattern to that in Figure 1.3 emerges, with U.S.-based producers appearing to source particular inputs quite differently depending on the location from which these products are bought. Imports from 17 of the 56 countries are exclusively transacted at arm's-length, while one country (Liechtenstein) sells steering wheels to the U.S. almost exclusively within multinational firm boundaries. The remaining 38 countries feature shares of intrafirm trade fairly uniformly distributed between 0 and 100 percent.

The large variation in the relative importance of intrafirm transactions across types of goods and countries might seem to validate the skeptics' view that the decision to integrate or outsource foreign production processes is largely driven by idiosyncratic factors that cannot possibly be captured by

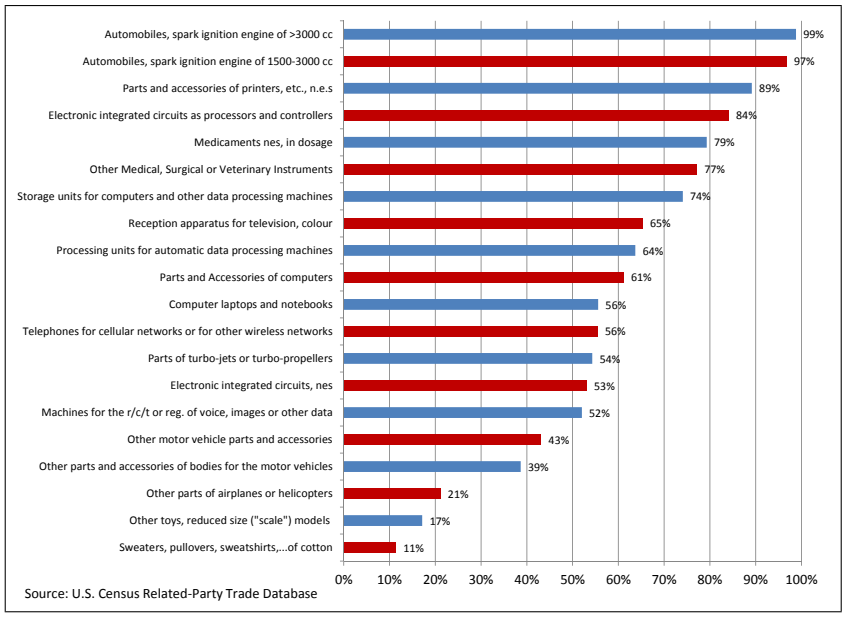

Figure 1.4: Share of U.S. Intrafirm Imports in Top 20 Industries by U.S. Import Volume in 2011

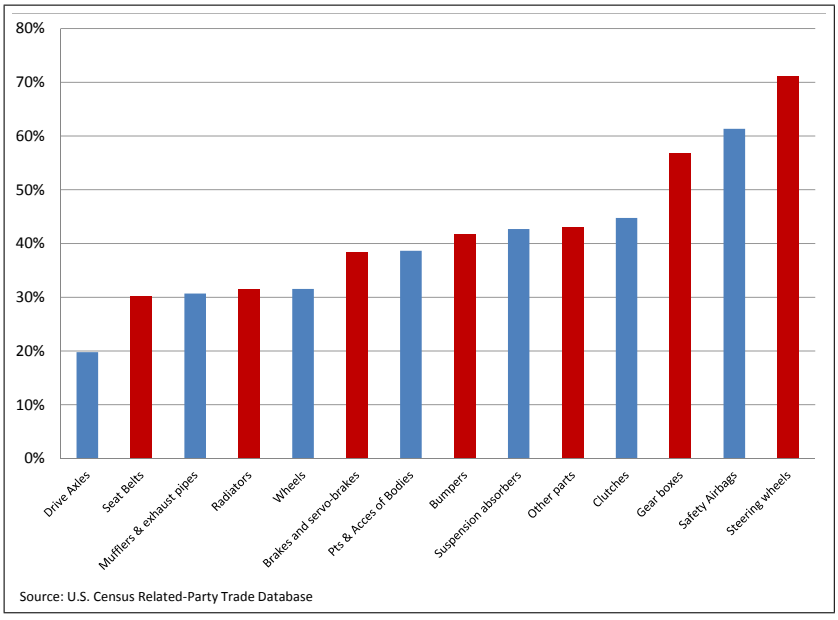

Figure 1.5: Variation in the Share of Intrafirm Trade within HS Sector 8708 (Auto Parts) in 2011

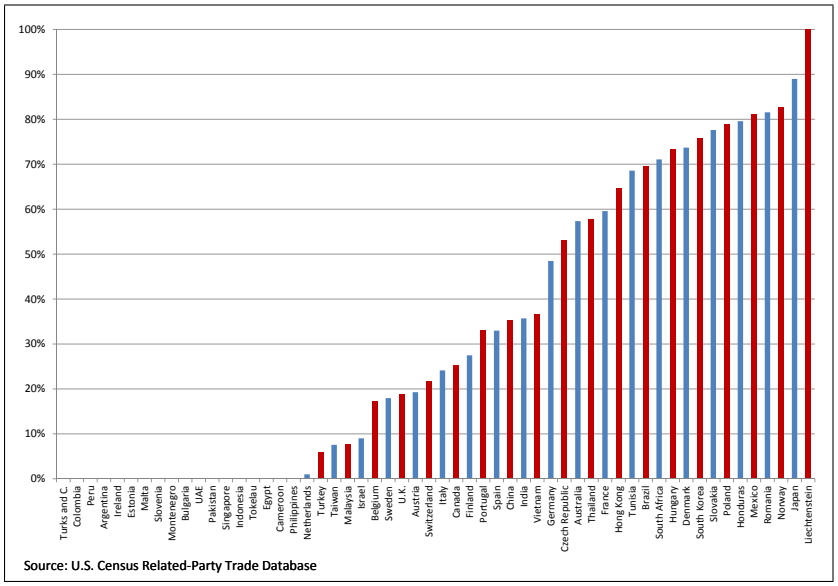

Figure 1.6: Variation in the Share of U.S. Intrafirm Imports within HS 870884 (Steering Wheels) in 2011

parsimonious models of the organization decisions of firms. If that were the case, however, not only would we observe large variation in the share of intrafirm trade, but we would also expect this variation to be uncorrelated with simple industry or country-level variables. As first demonstrated by Antràs (2003), the evidence suggests otherwise. Chapter 8 will describe in detail several stylized facts regarding the intrafirm component of trade. As a sneak preview, Figures 1.7 and 1.8 illustrate that the share of intrafirm imports in total U.S. imports is significantly higher, the higher the U.S. capital intensity in production of the good being imported and is also significantly higher, the higher the capital-labor ratio of the exporting country. These scatter plots suggest that, as argued above, there may indeed be some common fundamental factors that shape the integration decisions of firms across sectors and countries. The theories of internalization exposited in Chapters 6 and 7 will attempt to shed some light on these factors and will provide a valuable lens through which to study the intrafirm trade data in a more formal and structured manner.

While several features of the U.S. Related-Party Trade database make it particularly attractive to empirical researchers, it has some important limitations. Some of the shortcomings of the data relate to the extent to which the characteristics of the data permit a formal test of the theories of in-

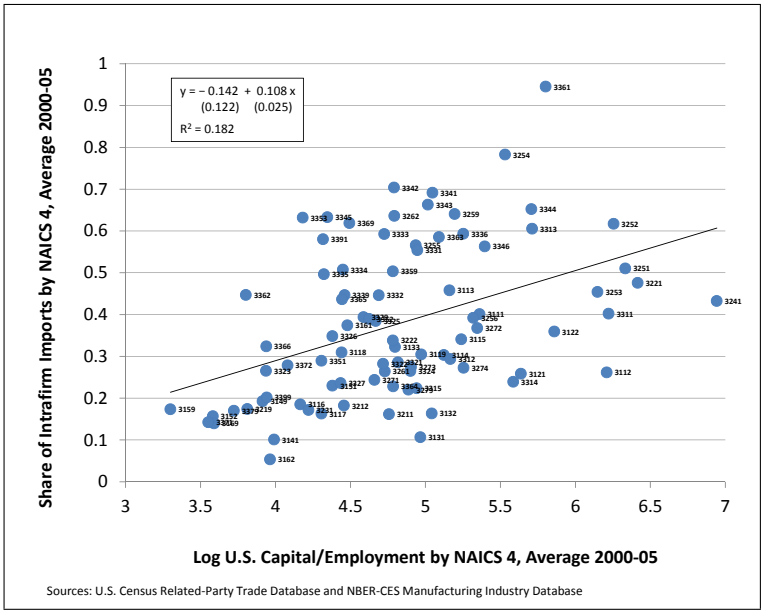

Figure 1.7: The Share of Intrafirm U. S. Imports and Capital Intensity

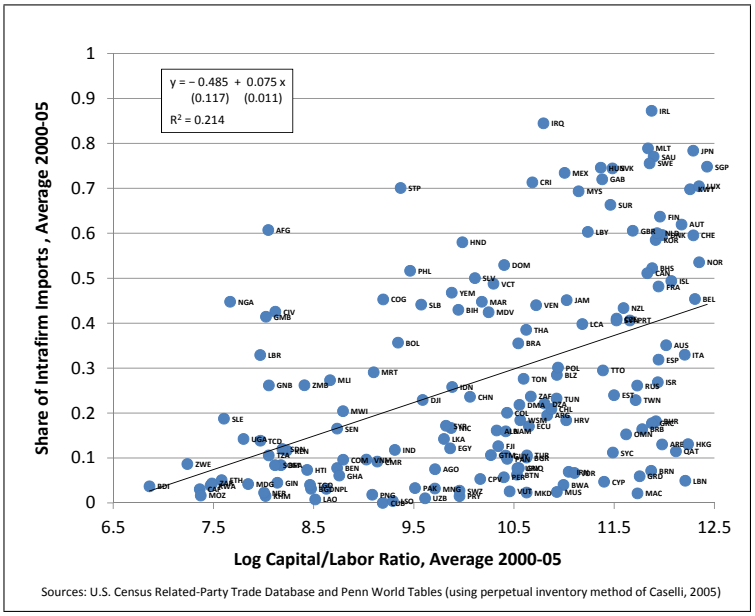

Figure 1.8: The Share of Intrafirm U. S. Imports and Capital Abundance

ternalization developed later in the book, so it is convenient to postpone that discussion until after we have covered those theories in Chapters 6 and 7. Other potential limitations are more fundamental, so it is important to tackle them upfront.

The U.S. database defines 'related-party imports' as import transactions between parties with various types of relationships including "any person directly or indirectly, owning, controlling or holding power to vote, 6 percent of the outstanding voting stock or shares of any organization." A first natural concern is then that the 6 percent threshold might be too low for that 'relatedness' to have any significant economic meaning, such as one of the entities having a *controlling* stake in the other entity. In practice, however, extracts from the confidential foreign direct investment dataset collected by the Bureau of Economic Analysis suggest that intrafirm trade is generally associated with one of the entities having a majority-ownership stake in the other entity. More specifically, in 2009, of all U.S. imports associated with U.S. parents purchasing goods from their affiliates, 93.8% involved majority-owned foreign affiliates. Similarly, majority-owned U.S. affiliates accounted for 95.5% of U.S. imports by all U.S. affiliates of foreign companies in 2009.[21]

A second general concern relates to overall quality of the data. In that respect, the technical documentation that accompanies the dataset stresses that the data are not subject to sampling error, since an indicator of whether the transaction involves related parties or not is required for *all* import or export transactions recorded by the U.S. Bureau of Customs and Border Protection. Despite this requirement, importers and exporters do not always report that information in their shipment documents. Luckily, these transactions are categorized on the data tables as "nonreported," so it is easily verified that these account for a very low share of trade volumes (for instance, just 1.4 percent of total imports in 2011). One might also worry about non-sampling errors related to the imputation of trade values for undocumented shipments and for low-valued transactions (which are sometimes estimated). Nevertheless, quality assurance procedures are performed at every stage of collection, processing, and tabulation, thus there is no reason to believe that these data are any less reliable than U.S. customs data on trade flows.[22]

One way to gain reassurance regarding the usefulness of the data is to see whether it delivers patterns that are consistent with what one would expect based on independent and reliable sources of data. For example, from

[21] See Table 9 in http://www.bea.gov/scb/pdf/2011/11%20November/1111_mnc.pdf, and Table I.A.1 in http://www.bea.gov/international/pdf/fdius_2009p/I%20A1%20to%20I%20A9.pdf.

[22] Ruhl (2013) provides a useful overview of alternative U.S. intrafirm trade data sources.

a quick search of press releases from recent years, one learns that in 2005, Boston-based Gillette Company completed the construction of a 120 million-euro plant in Łódź (Poland), which manufactures disposable razors and other shaving products.[23] Although production was mostly directed to the European market, it seems reasonable to assume that some of the products produced in the plant were shipped back to the U.S., a transaction that would naturally occur within firm boundaries. As shown in Figure 1.9, it is reassuring to observe that the share of intrafirm imports in total U.S. imports from Poland of NAICS code 332211, which is dominated by non-electric razors and razor blades, went up dramatically around the time of the plant opening, jumping from essentially 0 percent in 2004 to close to 100 percent from 2005 onwards.

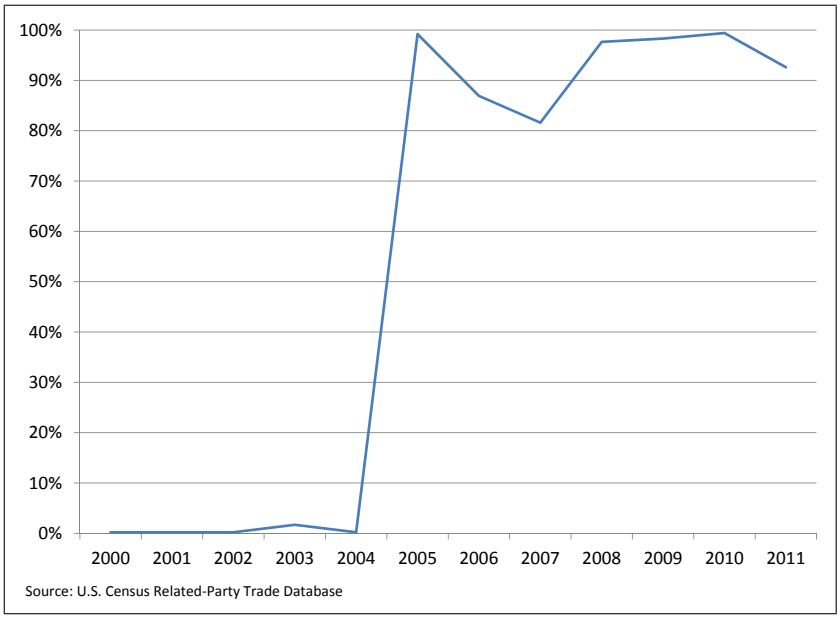

Figure 1.9: Share of Intrafirm Imports of NAICS 332211 (Razors) from Poland

Back to the Location Decision

We have emphasized above that the internalization decisions of firms in the global economy cannot be understood without appealing to contractual frictions and we have also illustrated the importance of these frictions in the

[23] See http://www.paiz.gov.pl/nowosci/?id_news=502.

real world. It seems natural, however, to posit that imperfect contracting not only shapes the ownership structure decisions of firms but might also impact their geographical location decisions. As emphasized by neoclassical models of offshoring, profit-maximizing firms will organize production in a cost-minimizing manner, but the effective costs of doing international business are not solely explained by the factors highlighted by neoclassical theory. Certainly, wages will, other things equal, tend to be relatively lower in relatively labor-abundant countries. And, other things equal, costs of production will also tend to be relatively low in countries or regions where the technologies used in production are particularly advanced. Yet, firms might be reluctant to offshore production lines to low-wage countries where suppliers are unreliable and tend not to honor their contracts, and where local courts are unlikely to effectively enforce contracts. Similarly, firms might be unwilling to operate in countries in which their advanced technologies could be effectively deployed (given the existence of local complementary factors), but in which the contractual environment might not provide enough security to firms, both in terms of quality contracting but also in terms of the risk of intellectual property rights expropriation.

A key factor that makes contractual aspects important for sourcing decisions is the existence of huge variation among countries in judicial quality and contract enforcement. Empirical researchers often make use of easily accessible measures of the quality of the rule of law which are themselves based on weighted averages of various indices of the perceived effectiveness and predictability of courts in different countries. An advantage to these widely used measures, such as the 'Rule of Law' variable produced by the Worldwide Governance Indicators, is that they capture broad features of the contracting environment. A disadvantage is that they are partly based on subjective assessments rather than objective measures of institutional quality. Furthermore, they may provide a useful ordinal measure of legal quality but they are less well equipped to help quantify the existence of cross-country heterogeneity in judicial quality and contract enforcement.

Djankov, La Porta, Lopez-De-Silanes and Shleifer (2003) have proposed an ingenious alternative measure of judicial quality which is narrower in nature but more powerful in illustrating the relevance of differences in the legal system across countries. In particular, Djankov al. (2003) estimate for 109 countries the time it takes a plaintiff using an official court to evict a nonpaying tenant and to collect a bounced check. Figure 1.10 depicts the second of these two variables, which is more likely to be of relevance for firms considering doing business in a particular country. Their estimated total duration of a legal procedure aimed at collecting a bounced check ranges

from 7 days in Tunisia to 1003 in Slovenia. Even when focusing on the 43 of the top 50 largest exporters to the U.S. for which they provide data, the estimated duration ranges from 39 days for the Netherlands to 645 days for Italy.

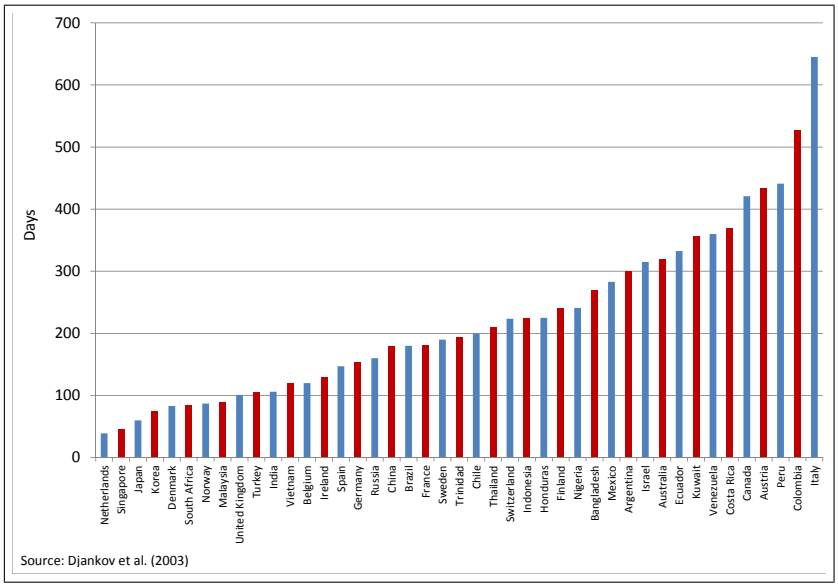

Figure 1.10: Duration of a Legal Procedure Aimed at Collecting a Bounced Check

The extent of contractual insecurity not only varies across countries (or jurisdictions) but it naturally also varies depending on the characteristics of the goods being transacted. For instance, basic goods with low levels of differentiation and which are traded in relatively thick markets can be relatively safely procured even from countries with weak contracting institutions. Conversely, transactions involving highly complex or differentiated goods will tend to be much more 'contract dependent', and one would expect firms to be significantly more sensitive to the institutional environment when choosing the country from which to procure those goods.

A Brief Road Map

This book will study the various ways in which the contracting environment shapes the location and internalization decisions of firms in the global economy. I will focus first on an analysis of the location decision and how it is affected by contracting factors, and only in Part III of the book will I allow firms to optimally decide the extent of control they want to exert over

production processes. This does not follow the chronological order in which these topics were developed in the literature, but I will adopt this sequencing for pedagogical reasons.

Before diving into the world of incomplete contracts, it is necessary, however, to provide an overview of the 'complete-contracting' frameworks that will serve as the basis or skeleton for the models to be developed in future chapters. A succinct overview of these models is offered in Chapter 2, to which I turn next. Readers familiar with Melitz' (2003) classic paper and its various extensions might want to jump straight to Part II of the book, starting in Chapter 3.